ACML 2017

Asian Conference on Machine Learning

The 9th Asian Conference on Machine Learning

November 15 - 17, 2017

Yonsei University, Seoul, Korea

Conference Book

Contents

Welcome Message from General Co-Chairs

On behalf of the organizing committee, it is our great pleasure to welcome you to Seoul, Korea for the 9th Asian Conference on Machine Learning (ACML 2017). ACML is a premier conference on machine learning providing a very friendly environment for the leading researchers in the Asia-Pacific region and beyond. The conference this year is jointly organized by the Yonsei University, Korea Advanced Institute of Science and Technology (KAIST), and Korea Robotics Society. We have prepared a three-day program fully packed with keynote and invited talks, tutorials, workshops, and the main track technical paper presentations.

We are excited to hold the conference at the brand-new conference facility at Yonsei University, which is at the heart of Sinchon area full of creativity and energy built by young generations. We hope that you enjoy the vibrant culture and local food in the neighborhood throughout the conference.

The conference would not have been possible without the efforts of many people, most importantly the organizing committee: Program Co-Chairs Yung-Kyun Noh and Min-Ling Zhang, Journal Track Co-Chairs Wee Sun Lee and Bob Durrant, Workshop Co-Chairs Krikamol Maundet and Jihum Hamm, Tutorial Co-Chairs Hung Bui and Jaesik Choi, and Publicity Co-Chairs Hsuan-Tien Lina and Ivor Tsang. We are also greatly thankful to Honorary Co-Chairs Hyeran Byun and Byoung-Tak Zhang for their leadership guidance. Our thanks also go to Local Arrangements Chair Seon Joo Kim and Conference Secretary Hyangmi Kim who did a great job on handling logistics problems that turned out to be much more complex than initially expected. We also appreciate the ACML Steering Committee for their decision on accepting Seoul, Korea to host ACML 2017 for the first time, and sponsoring best paper awards and student travel grants.

We are grateful for the sponsors and supporters, whose contributions were huge for the success of the conference. Our thanks go out to Yonsei University, Korea Advanced Institute of Science and Technology (KAIST), Big Data Institute of Seoul National University, RIKEN Center for Advanced Intelligence Project, AFOSR/AOARD, Samsung SDS, Naver, NVIDIA, Amazon AWS, Korea Tourism Organization, and Seoul Metropolitan Government.

Last but not least, we would like to thank all the participants and attendees who define the conference this year.

Welcome and enjoy the conference and have a great time in Korea!

Kee-Eung Kim
KAIST, Korea
General Co-Chair

Masashi Sugiyama
RIKEN/ The University of Tokyo, Japan
General Co-Chair

Welcome Message from Program Co-Chairs

Welcome to the Ninth Asian Conference on Machine Learning (ACML 2017) in Seoul, Korea. The ACML continues the tradition of having high-quality and original research papers in the area of machine learning following eight previous successful events held in China, Japan, Taiwan, Singapore, Australia, Vietnam, Hong Kong, and New Zealand respectively. ACML aims at providing a leading international forum for researchers in machine learning and related fields to share their original research findings, new ideas and achievements. Despite originating in the Asia-Pacific region, ACML has become a worldwide conference: This year accepted papers are based in Canada, Germany, France, Finland, Netherland, India, and the Northeast of the USA, as well as the Asia-Pacific region.

This year, there were 172 submissions for the two cycles of the conference track. A strict double-blind reviewing process was enforced, and each paper was assigned with one Senior Program Committee member and at least three Program Committee members, who provided expert opinions and contributed with discussions after author response to their reviews. Finally, 41 are accepted into the main program, for an acceptance rate of 23.8%. Those accepted proceedings are published in volume 77 of Proceedings of Machine Learning Research. Following the new innovation of last year, this year ACML also ran an additional journal track. The journal track Co-Chairs oversee the reviewing process of 23 submissions, out of which 6 papers are selected for publication in the Springer journal Machine Learning, for an acceptance rate of 26%.

For ACML this year the overall number of accepted papers, from both the journal and proceedings tracks, was 47 from 195 submissions for a 24.1% total acceptance rate.

The accepted papers from the two tracks cover a broad range of topics, including theoretical analyses, probabilistic models, large-scale machine learning, weakly-supervised/unsupervised learning, multi-view/multi-task/crowdsourced learning, deep learning, and applications to real world problems. All participants of ACML 2017 are welcome to joining these presentations and hope enjoy the active discussions among participants.

November 2017
ACML 2017 Program Co-Chairs

Yung-Kyun Noh
Seoul National University, Korea

Min-Ling Zhang
Southeast University, China

Organizing Committee

Honorary Chairs

Hyeran Byun	Yonsei University, Korea
Byoung-Tak Zhang	Seoul National University, Korea

General Co-Chairs

Kee-Eung Kim	KAIST, Korea
Masashi Sugiyama	RIKEN/ The University of Tokyo, Japan

Program Co-Chairs

Yung-Kyun Noh	Seoul National University, Korea
Min-Ling Zhang	Southeast University, China

Journal Track Co-Chairs

Wee Sun Lee	National University of Singapore, Singapore
Bob Durrant	University of Waikato, New Zealand

Workshop Co-Chairs

Krikamol Muandet	Mahidol University / Max Plank Institute, Thailand/ Germany
Jihun Hamm	Ohio State University, USA

Tutorial Co-Chairs

Hung Bui	Adobe Research, USA
Jaesik Choi	UNIST, Korea

Publicity Co-Chairs

Hsuan-Tien Lin	National Taiwan University, Taiwan
Ivor Tsang	University of Technology Sydney, Australia

Local Arrangements Chair

Seon Joo Kim	Yonsei University, Korea

Senior Program Committee

Wray Buntine	Monash University, Australia
Marco Cuturi	ENSAE/CREST, France
Stephen Gould	Australian National Univ., Australia
Yuhong Guo	Temple University, USA
Steven Hoi	Singapore Management Univ,, Singapore
Sheng-Jun Huang	Nanjing Univ. of Aeronautics and Astronautics, China
Tomoharu Iwata	NTT Communication Science Laboratories, Japan
Junmo Kim	KAIST, Korea
Young Jun Ko	EPFL, Switzerland
James Kwok	Hong Kong University of Science and Technology, Hong Kong China
Hoai An Le Thi	University of Metz, France
Yuh-Jye Lee	National Taiwan University of Science and Technology, Taiwan
Sangkyun Lee	Hanyang University, Korea

Yu-Feng Li	Nanjing University, China
Hsuan-Tien Lin	National Taiwan University, Taiwan
Shoude Lin	National Taiwan University, Taiwan
Krikamol Muandet	Mahidol University / Max Plank Institute, Thailand/ Germany
Alice Oh	KAIST, Korea
Sinno Pan	Nanyang Technological University, Singapore
Bernhard Pfahringer	University of Waikato, New Zealand
Dinh Phung	Deakin University, Australia
Tao Qin	Microsoft Research Asia, China
Jesse Read	Université Paris-Saclay, France
Truyen Tran	Deakin University, Australia
Grigorious Tsoumakas	Aristotle University of Thessaloniki, Greece
Liwei Wang	Peking University, China
Jun Xu	Institute of Computing Technology, CAS, China
Florian Yger	Université Paris-Dauphine, France
Yang Yu	Nanjing University, China
Junping Zhang	Fudan University, China
Kun Zhang	CMU, USA
Jun Zhu	Tsinghua University, China

Journal Track Committee

Wray Buntine	Monash University, Australia
Steven Hoi	Singapore Management University, Singapore
Kee-Eung Kim	KAIST, Korea
Hang Li	Toutiao, China
Yu-Feng Li	Nanjing University, China
Masashi Sugiyama	RIKEN/ University of Tokyo, Japan
Ichiro Takeuchi	Nagoya Institute of Technology, Japan
Ivor Tsang	University of Technology Sydney, Australia
Lijun Zhang	Nanjing University, China
Jun Zhu	Tsinghua University, China

Steering Committee

Zhi-Hua Zhou	Nanjing University, China (Chair)
Masashi Sugiyama	RIKEN/ The University of Tokyo, Japan (Co-Chair)
Thomas G. Dietterich	Oregon State University, USA
Tu-Bao Ho	JAIST, Japan
Wee Sun Lee	National University of Singapore, Singapore
Wray Buntine	Monash University, Australia
Chi-Jen Lin	National Taiwan University, Taiwan
Hiroshi Motoda	AFOSR/AOARD and Osaka University, Japan

Sponsored by

Sponsored by

Yonsei University

KAIST College of Engineering / International Relations Team

Seoul National University / Big Data Institute

Korea Robotics Society

Supported by

Google

Samsung SDS

RIKEN Center for Advanced Intelligence Project

Amazon AWS

Naver

NVIDIA

AFOSR/AOARD, ITC-PAC

GIGA-BYTE Technology

Korea Tourism Organization

Seoul Metropolitan Government

Exhibitors

Conference Venue

ACML2017 will be held at Baekyang-Nuri, Yonsei University in Seoul, Korea. Yonsei University is one of the most prestigious universities in Korea. It began as a medical institute created by Western missionaries about 130 years ago. Today, Yonsei takes great pride in its over 300,000 alumni, 4,800 faculty members and four campuses including Sinchon, Wonju and Songdo. There are over 25,000 undergraduates and graduate students on Sinchon campus. Many of its alumni have achieved leadership positions in Korea. Baekyang-Nuri was renovated in October, 2015, to provide faculty and students with state of the art facilities and relaxation spaces. It will serve as an excellent venue for conferences such as ACML.

Venue Address

Baekyang-Nuri (Building #130), Yonsei University
Address: 50 Yonsei-ro, Seodaemun-gu, Seoul 03722, Korea
Website: http://www.yonsei.ac.kr/
Phone: 1599-1885

Location of Conference Venue

Baekyang-Nuri (building #130) is at 100 meters distance from the main gate on your left side. Get off at "Sinchon" station (subway line #2) and get out to way #2 or #3. Go straight toward "Yonsei-ro" and find out the main gate. Refer to the campus map on the right side.

How to get to Yonsei University from Incheon Airport

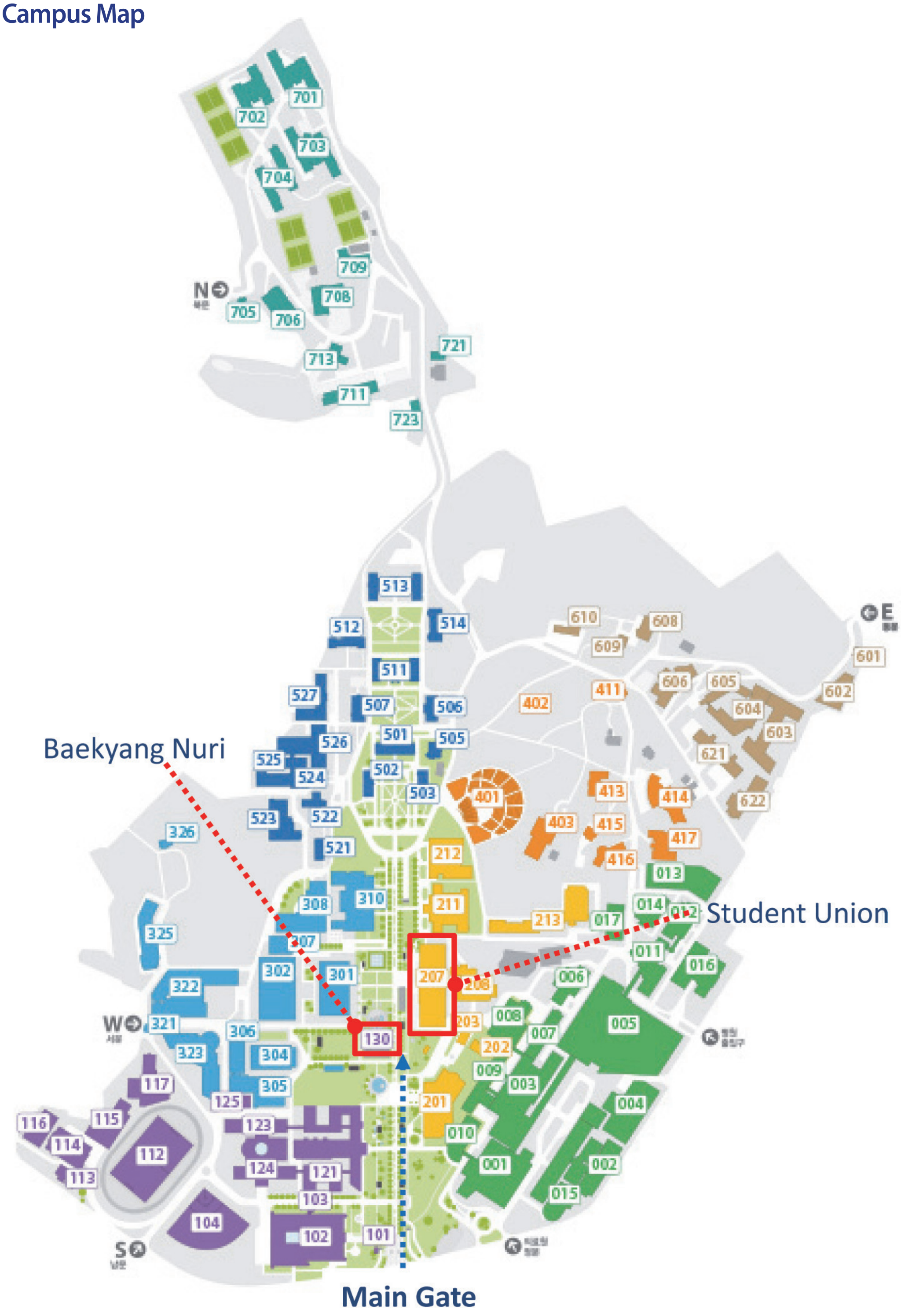
• Airport Limousine Bus

Take a limousine bus #6011 or #6002 at bus stop #5B or #11B on the 1st floor of the Incheon International Airport. Bus #6011 gets you off at the main gate of Yonsei University, while Bus #6002 drops you at Sinchon subway station. Duration is about one hour. Bus fare is 10,000 Won. It operates every 20 minutes from 5 am to 8:30 pm.

• Subway

Take a train, Airport Railroad (AREX) and transfer to Line #2 at “Hongik University” station. Get off at "Sinchon" station (subway line #2) and get out to way #2 or #3. Duration is 60 minutes. Fare is about 4200 Won.

Yonsei Campus Map



Registration

The registration desk will be open on the lobby of Baekyang-Nuri (Building #130) as follows:

Wednesday, 15 November	8 am - 6 pm
Thursday, 16 November	8 am - 6 pm
Friday, 17 November	8 am - 5 pm

Pre-registrants can pick up a registration kit there. The conference registration kit includes a conference bag, conference book, a copy of digital proceedings, name badge, receipt of registration fee, three lunch coupons and a banquet coupon. All attendees must wear their name badges at all times to gain admission to all conference sessions, and to the reception, lunch and banquet. Tutorial & Workshop fee includes one lunch, refreshment break, and conference book.

Presentations, Equipment, and Wireless Internet Access

All rooms assigned for oral presentations, workshops and tutorials are equipped with an LCD projector with VGA connector and screen. During the single track of the conference, each paper will be presented in both oral and poster form.

Oral Presentations will be held in Grand Ballroom with one session on November 15 afternoon, three sessions on November 16, and the remaining three sessions on November 17. The time slot assigned for each paper in the session is 17 minutes, including 15 minutes for oral presentation and 2 minutes for Q/A. Speakers should bring a laptop with their presentation slides ready and be prepared to display over a VGA cable. Please note that the power outlets are mostly 220 volts and 60 Hertz in Korea.

Poster Presentations will take place on the Lobby during lunch time. The poster board is 95 cm wide X 165 cm high. The actual space for poster material is 95 cm wide X 120 cm high. Poster material must be in place by 11 am before the poster session. It will remain in place until the end of the poster session. Stationery will be provided in the registration desk. All materials must be cleared out by 2 pm on the same date. Speakers are responsible for handling their poster. All posters left after 6 pm will be discarded.

Wireless internet access is available during the conference.

Wednesday, 15 November

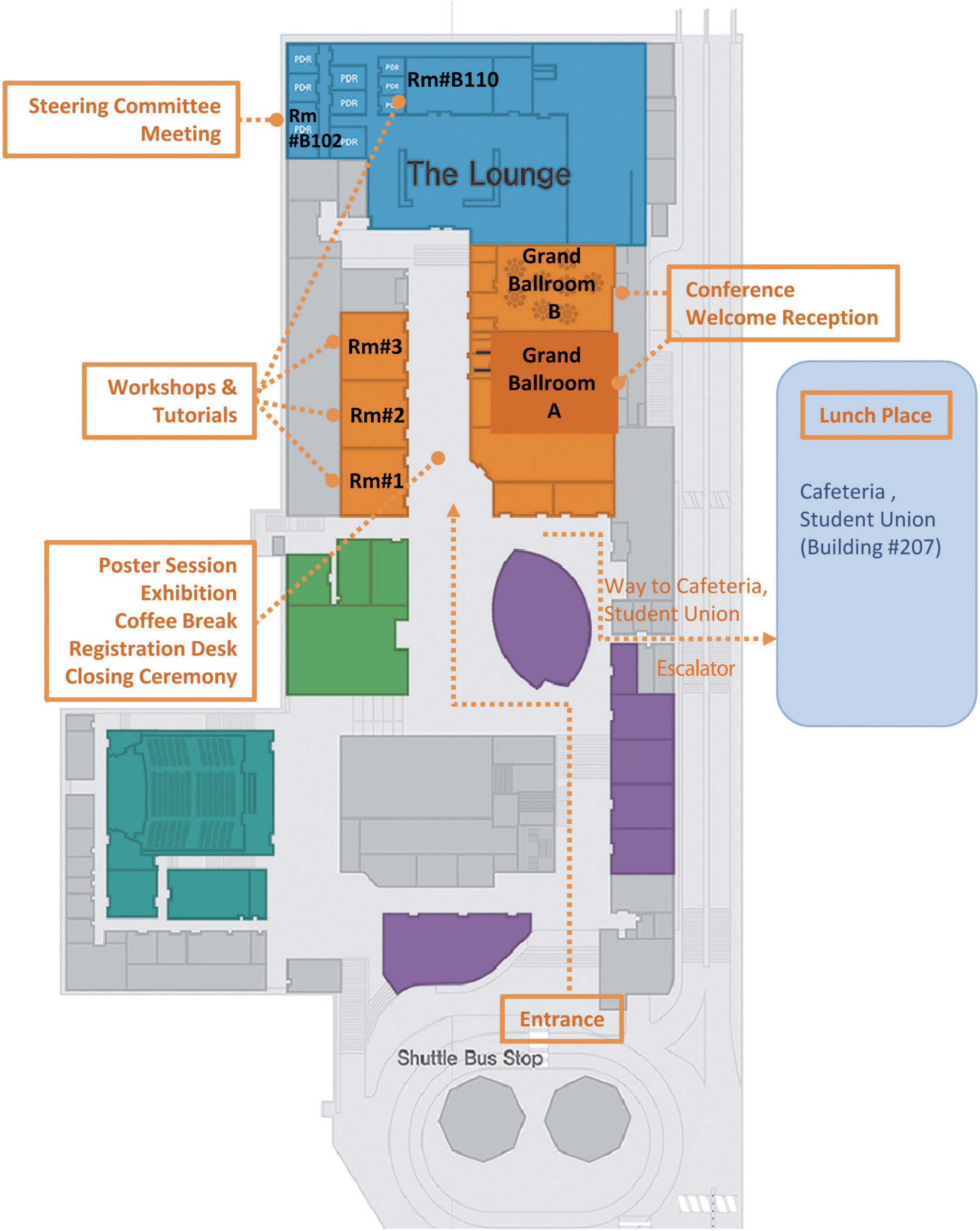
Time						
8:00	Registration @ Lobby					
8:30–11:30	Tutorial 1 @ RM#1 8:30-10:00 Tutorial 2 10:00-11:30	Tutorial 4 @ RM#2 8:30-11:30	Workshop 1 @ RM#3 8:30-15:30	Workshop 2 @ B110 8:30-15:30	Workshop 3 @ Grand Ballroom A 10:00-15:30	Exhibition @ Lobby
11:30 12:30	Lunch @ Cafeteria, Student Union 11:30 - 12:30					
12:30–15:30	Tutorial 3 @ RM#1 12:30-15:30	Tutorial 5 @ RM#2 12:30-15:30	Workshop 1 @ RM#3	Workshop 2 @ B110	Workshop 3 @ Grand Ballroom A	
	Break 15:30 - 16:00					
16:00–18:00	Oral Session 1 @ Grand Ballroom A Large-Scale Machine Learning 16:00 - 18:00					
18:30–20:00	Welcome Reception @ Grand Ballroom B 18:30 - 20:00					

Thursday, 16 November

Time		
8:20	Opening Ceremony 08:20-08:30	
8:30	Invited Talk 1 Speaker: Eunho Yang @ Grand Ballroom A+B 08:30 - 09:15	
	Break 09:15 - 09:35	
11:35	Oral Session 2 @ Grand Ballroom Statistical/Bayesian Machine Learning 09:35 - 11:35	Exhibition @ Lobby
11:35 13:00	Lunch @ Cafeteria, Student Union Poster Session @ Lobby 11:35 - 13:00	
13:00	Keynote Talk I Speaker: Bernhard Schölkopf @ Grand Ballroom 13:00 - 14:00	
16:00	Oral Session 3 @ Grand Ballroom Machine Learning Applications 14:00 - 16:00	
	Break 16:00 - 16:20	
16:20–18:10	Oral Session 4 @ Grand Ballroom Weakly-supervised/Unsupervised Learning 16:20 - 18:10	
18:20–21:00	Conference Banquet @ Fradia 19:00 - 21:00	

Friday, 17 November

Time		
8:30	Inivited Talk 2 Speaker: Hang Li @ Grand Ballroom A+B 08:30 - 09:15	
	Break 09:15 - 09:35	
11:35	Oral Session 5 @ Grand Ballroom Deep Learning 09:35 - 11:35	Exhibition @ Lobby
11:35 13:00	Lunch @ Cafeteria, Student Union Poster Session @ Lobby 11:35 - 13:00	
13:00	Keynote Talk II Speaker: Tom Dietterich @ Grand Ballroom 13:00 - 14:00	
15:50	Oral Session 6 @ Grand Ballroom Learning Theory 14:00 - 15:50	
	Break 15:50 - 16:10	
16:10–18:10	Oral Session 7 @ Grand Ballroom Multi-view, Multi-task, and Crowdsouced Learning 16:10 - 18:10	
18:20–19:00	Closing Ceremony @ Lobby 18:20 - 19:00	


Steering Committee Meeting
Rm #B102
Rm#B110
PDR
The Lounge
Grand Ballroom B
Conference Welcome Reception
Grand Ballroom A
Workshops & Tutorials
Rm#3
Rm#2
Rm#1
Lunch Place
Cafeteria , Student Union (Building #207)
Poster Session
Exhibition
Coffee Break
Registration Desk
Closing Ceremony
Way to Cafeteria, Student Union
Escalator
Entrance
Shuttle Bus Stop

Keynote Talks

Invited Talks

Tutorials

Workshops

Conference on Wedneday, 15 November
Oral Session

Conference on Thursday, 16 November
Oral Session / Poster Session

Conference on Friday, 17 November
Oral Session / Poster Session

ACML 2017
Asian Conference on Machine Learning

Chair: Masashi Sugiyama

"Causal Learning"

Bernhard Schölkopf
Professor and Director of Max Planck Institute for Intelligent Systems, Germany

Abstract: In machine learning, we use data to automatically find dependences in the world, with the goal of predicting future observations. Most machine learning methods build on statistics, but one can also try to go beyond this, assaying causal structures underlying statistical dependences. Can such causal knowledge help prediction in machine learning tasks? We argue that this is indeed the case, due to the fact that causal models are more robust to changes that occur in real world datasets. We discuss implications of causality for machine learning tasks, and argue that many of the hard issues benefit from the causal viewpoint. This includes domain adaptation, semi-supervised learning, transfer, life-long learning, and fairness, as well as an application to the removal of systematic errors in astronomical problems.

Biography: Bernhard Schölkopf's scientific interests are in machine learning and causal inference. He has applied his methods to a number of different application areas, ranging from biomedical problems to computational photography and astronomy. Bernhard has researched at AT&T Bell Labs, at GMD FIRST, Berlin, and at Microsoft Research Cambridge, UK, before becoming a Max Planck director in 2001. He is a member of the German Academy of Sciences (Leopoldina), and has received the J.K. Aggarwal Prize of the International Association for Pattern Recognition, the Max Planck Research Award (shared with S. Thrun), the Academy Prize of the Berlin-Brandenburg Academy of Sciences and Humanities, and the Royal Society Milner Award.

Chair: Kee-Eung Kim

"Combining AI and Visualization to Manage Ecosystems"

Tom Dietterich, Distinguished Professor (Emeritus) and Director of Intelligent Systems, Oregon State University, USA

Abstract: As humans occupy virtually all of the planet, we must actively manage ecosystems in order to ensure their sustained functioning. Many ecosystem problems can be formulated as Markov Decision Problems (MDPs) in which the state transitions are provided by a simulator rather than by an explicit function. The two key technical challenges in solving such MDPs are (a) the state spaces are immense and (b) the simulators are very expensive. A third, more political, challenge (c) is that ecosystem management problems involve many stakeholders who often disagree about the objectives to be optimized. To address the first problem, we employ search in a parameterized policy space. We have obtained excellent results using SMAC, a form of Bayesian Optimization, to find good policies. To address the second problem, we extend the method of Model-Free Monte Carlo (MFMC) to create a surrogate model. We reduce the size of the state space by factoring out exogeneous state variables. To address the third problem, we have built a visualization environment, MDPvis, that allows multiple stakeholders to modify the MDP reward function and explore the behavior of the system. We hope that the combination of visualization and rapid MDP solution will help multiple stakeholders arrive at consensus on how to manage complex ecosystems.

Biography: Dr. Dietterich (AB Oberlin College 1977; MS University of Illinois 1979; PhD Stanford University 1984) is Professor Emeritus and Director of Intelligent Systems Research in the School of Electrical Engineering and Computer Science at Oregon State University, where he joined the faculty in 1985. Dietterich is one of the pioneers of the field of Machine Learning and has authored more than 130 refereed publications and two books. His research is motivated by challenging real world problems with a special focus on ecological science, ecosystem management, and sustainable development. He is best known for his work on ensemble methods in machine learning including the development of error-correcting output coding. Dietterich has also invented the MAXQ decomposition for hierarchical reinforcement learning.

Dietterich has devoted many years of service to the research community. He is Past President of the Association for the Advancement of Artificial Intelligence, and he previously served as the founding president of the International Machine Learning Society. Other major roles include Executive Editor of the journal Machine Learning, co-founder of the Journal for Machine Learning Research, and program chair of AAAI 1990 and NIPS 2000. Dietterich is a Fellow of the ACM, AAAI, and AAAS.ces and Humanities, and the Royal Society Milner Award.

Invited Talk 1 | **Grand Ballroom** | Thursday, 16 November, 8:30 am – 9:15 am

Chair: Yung-Kyun Noh

"Beyond Gaussian/Ising Graphical Models"

Eunho Yang, Assistant Professor in the School of Computing at KAIST, Korea

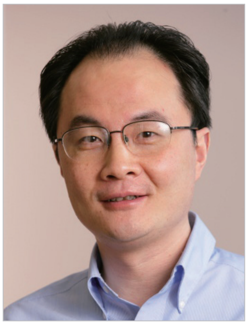
Abstract: Graphical models are the standard toolkit to model interactions between huge number of multiple random variables. Gaussian graphical models for continuous data and Ising (or discrete graphical models) for discrete data are the two most popular instances of undirected graphical models. In this talk, we will discuss limitations of these popular instances of pairwise graphical models with respect to two orthogonal directions: i) restriction on types of data following bell-shpaed Gaussian or categorical properties ii) lack of higher-order interactions due to computational overheads. We will then introduce more general graphical models beyond Gaussian and Ising graphical models, to overcome these limitations.

Biography: Eunho Yang is an assistant professor at the School of Computing, KAIST. Before joining KAIST, he spent two years at IBM T.J. Watson Research Center as a Research Staff Member. He obtained his Ph.D. in 2014 from the university of Texas at Austin, and did M.S. and B.S from the Seoul National University, Korea in 2006 and 2004, respectively. His research interests are in statistical machine learning in general with the special focuses on high-dimensional statistics. He is currently developing new theories and algorithms for graphical models and deep learning, with the applications of computational biology and medicine, etc.

Chair: Min-Ling Zhang

"Beyond Deep Learning: Combining Neural Processing and Symbolic Processing"

Hang Li, Toutiao, China



Abstract: Recently deep learning has brought significant breakthroughs to natural language processing. I will start the talk by summarizing the strengths and limitations of deep learning in natural language processing. I will then indicate that a hybrid approach combining neural processing (deep learning) and symbolic processing would be necessary and even more powerful for the tasks in natural language processing, particularly question answering. This is because the two paradigms are both advantageous and complementary; symbolic representations are easier to interpret and manipulate, while neural representations are more suitable for dealing with uncertainty in language and noise in data. Finally, I will introduce our recent effort toward developing neural symbolic processing, including building of a hybrid system for question answering from relational database and a hybrid system for generative question answering from knowledge base.

Biography: Hang Li is director of Toutiao AI Lab, adjunct professors of Peking University and Nanjing University. He is an IEEE Fellow and an ACM Distinguished Scientist. His research areas include information retrieval, natural language processing, machine learning, and data mining. Hang graduated from Kyoto University in 1988 and earned his PhD from the University of Tokyo in 1998. He worked at NEC Research as researcher from 1990 to 2001, Microsoft Research Asia as senior researcher and research manager from 2001 to 2012, and chief scientist and director of Huawei Noah's Ark from 2012 to 2017. He joined Toutiao in 2017. Hang has published three technical books, and more than 120 technical papers at top international conferences including SIGIR, WWW, WSDM, ACL, EMNLP, ICML, NIPS, SIGKDD, AAAI, IJCAI, and top international journals including CL, NLE, JMLR, TOIS, IRJ, IPM, TKDE, TWEB, TIST. He and his colleagues' papers received the SIGKDD'08 best application paper award, the SIGIR'08 best student paper award, the ACL'12 best student paper award. Hang worked on the development of several products such as Microsoft SQL Server 2005, Office 2007, Live Search 2008, Bing 2009, Office 2010, Bing 2010, Office 2012, Huawei smartphones 2014 and Huawei smartphones 2017. He has 42 granted US patents. Hang is also very active in the research communities and has served or is serving top international conferences as PC chair, Senior PC member, or PC member, including SIGIR, WWW, WSDM, ACL, NACL, EMNLP, NIPS, SIGKDD, ICDM, IJCAI, ACML, and top international journals as associate editor or editorial board member, including CL, IRJ, TIST, JASIST, JCST.

High Dimensional Causation Analysis

Zhenjie Zhang, Advanced Digital Sciences Center, Singapore

Ruichu Cai, Guangdong University of Technology, China

Causation analysis is one of the most fundamental research topics in machine learning, which aims to identify causal variables linked to the effect variables from a group of sample in the high dimensional space. The result of causation analysis provides the key insight into the target problem domain, and potentially enable new technologies of genetic therapy in genomic domain and predictive maintenance in IoT domain. Different from existing regression and classification algorithms in machine learning by exploiting correlations among variables, e.g., random forest and deep learning, causation analysis is supposed to unveil the complete and accurate structure of causal influence between every pair of variables in the domain.

It raises extremely large challenges to both mathematical model and algorithm design, because of the exponential complexity growth by extending from correlational dependency to causal dependency. In last decade, huge efforts are devoted to a variety of research frontiers of causation analysis, generating interesting and impressive new understandings under completely different assumptions behind the underlying causal structure generation process. In this tutorial, we introduce the theoretical discoveries on new models, review the significance and usefulness of the new approaches, discuss the applicability of new algorithms on real world applications, and address possible future research directions.

Statistical Relational Artificial Intelligence

Jaesik Choi, Ulsan National Institute of Science and Technology (UNIST), Korea

Hung Bui, Adobe Research, USA

An intelligent agent interacting with the real world will encounter individual people, courses, test results, drugs prescriptions, chairs, boxes, etc., and needs to reason about properties of these individuals and relations among them as well as cope with uncertainty.

Uncertainty has been studied in probability theory and graphical models, and relations have been studied in logic, in particular in the predicate calculus and its extensions. This book examines the foundations of combining logic and probability into what are called relational probabilistic models. It introduces representations, inference, and learning techniques for probability, logic, and their combinations.

This tutorial will provide a gentle introduction into the foundations of statistical relational artificial intelligence, and will realize this by introducing the foundations of logic, of probability, of learning, and their respective combinations.

Deep learning for Biomedicine

Truyen Tran, Deakin University, Australia

The ancient board game of Go, once predicted to remain unsolved for decades, is no longer an AI challenge. Equipped with deep learning, the program AlphaGo of DeepMind beaten a human champion 4 to 1. Indeed, deep learning has enjoyed many record-breaking successes in vision, speech and NLP and has helped boost a huge interest in AI from both academia and industry. Perhaps the next most important area for deep learning to conquer is biomedicine. With the obvious benefits to mankind and a huge industry, deep learning for biomedicine has recently attracted a great attention in both industry and academia. While we hold a great optimism for its success, biomedicine is new to deep learning and there are unique challenges yet to be addressed.

This tutorial consists of two parts. Part I briefly covers main ideas behind state-of-the-art deep learning theory and practice. Part II guides practitioners through designing deep architectures for biomedicine to best address the challenges, some unique to the field.

Distributed Convex Optimization

Jun Moon, Ulsan National Institute of Science and Technology (UNIST), Korea

Convex optimization is a special class of mathematical optimization, where both the objective function and the constraint sets are convex. A lot of optimization problems in engineering, economics and science can be formulated as convex problems, and there are various reliable and efficient algorithms to solve them.

In this tutorial, we consider convex optimization theory. The topics for this tutorial are as follows;

- Convex sets and functions
- Unconstrained convex optimization
- Constrained convex optimization
- Distributed convex optimization
- Dynamic programming and linear quadratic control

Machine Learning for Industrial Predictive Analytics

Evgeny Burnaev, Skolkovo Institute of Science and Technology, Russia

Maxim Panov, Skolkovo Institute of Science and Technology, Russia

Alexey Zaytsev, Skolkovo Institute of Science and Technology, Russia

Approximation problems (also known as regression problems) arise quite often in industrial design, and solutions of such problems are conventionally referred to as surrogate models. The most common application of surrogate modeling in engineering is in connection to engineering optimization for predictive analytics. Indeed, on the one hand, design optimization plays a central role in the industrial design process; on the other hand, a single optimization step typically requires the optimizer to create or refresh a model of the response function whose optimum is sought, to be able to come up with a reasonable next design candidate.

The surrogate models used in optimization range from simple local linear regression employed in the basic gradient-based optimization to complex global models employed in the so-called Surrogate-Based Optimization (SBO). Aside from optimization, surrogate modeling is used in dimension reduction, sensitivity analysis, and for visualization of response functions. In this tutorial we are going to highlight main issues on how to construct and apply surrogate models, describe both state-of-the-art techniques and a few novel approximation algorithms, demonstrate the efficiency of the surrogate modeling methodology on several industrial engineering problems.

The First International Workshop on Machine Learning for Artificial Intelligence Platforms (MLAIP)

Recently, several successful AI systems such as Amazon Alexa, Google Assistant, and NAVER X LINE Clova are developed based on AI-assistant platforms. These AI platforms contain several common technologies including speech recognition/synthesis, natural language understanding, image recognition, and dialog recommendation.

Building a successful MLAIP is a challenging mission because it requires a novel combination of heterogeneous machine learning models in a unified framework with efficient data processing. The goals of this workshop is to investigate and advance important topics in Machine Learning for AI Platforms (MLAIPs) further. In addition, we expect to provide the collaboration opportunities to researchers on ML theory on diverse application domains as well as industrial engineers.

Keynote Speaker

Masashi Sugiyama (University of Tokyo)

Invited Speakers

Kyomin Jung (Seoul National University)
Dit-Yan Yeung (Hong Kong University of Science and Technology)
Lucy Eunjeong Park (Papago, NAVER Corp.)
Jung-Woo Ha (Clova, NAVER Corp.)

Organizing Committee

Byoung-Tak Zhang (Seoul National University)
Sungroh Yoon (Seoul National University)
Dit-Yan Yeung (Hong Kong University of Science and Technology)
Sung Kim (Hong Kong University of Science and Technology)
Jaesik Choi (Ulsan National Institute of Science and Technology)
Jung-Woo Ha (Clova, NAVER Corp.)

The 2nd Asian Workshop on Reinforcement Learning (AWRL'17)

Keynote Speaker

Thomas G. Dietterich, Oregon State University, USA

Invited Speakers

Mohammad Ghavamzadeh, Google Deepmind, USA
Alessandro Lazaric, Facebook AI Resarch, France
Wee Sun Lee, National University of Singapore
Shie Mannor, Technion, Isarel
Bruno Scherrer, INRIA, France
Songhwai Oh, Seoul National University, South Korea

Motivation

The Asian Workshop on Reinforcement Learning (AWRL) focuses on both theoretical foundations, models, algorithms, and practical applications. In the last few years, we have seen a growing interest in RL of researchers from different research areas and industries. We invite reinforcement learning researchers and practitioners to participate in this world-class gathering. We intend to make this an exciting event for researchers and practitioners in RL worldwide, not only for the presentation of top quality papers, but also as a forum for the discussion of open problems, future research directions and application domains of RL.

AWRL 2017 (in conjunction with ACML 2017) will consist of keynote talks, contributed paper presentations, and discussion sessions spread over a one-day period.

Organizing Committee

Tao Qin, Microsoft Research Asia
Paul Weng, SYSU-CMU Joint Institute of Engineering, China
Yang Yu, Nanjing University, China
Zongzhang Zhang, Soochow University, China

2017 Annual Korea AI Society Meeting

Korea AI society is an organization with five special interest groups: SIG CVPR Pattern Recognition, SIG Machine Learning, SIG Artificial Intelligence, SIG Bio and Health, and SIG Brain Computer Interface. This is an annual joint meeting of five SIGs, introducing recent research achievements in each SIG. This meeting is also an opportunity for meeting researchers in working in different research areas for potential collaboration.

Keynote Speaker

Byoung-Tak Zhang, Seoul National University

Speakers

Sung Ju Hwang, Ulsan National Institute
Byungkon Kang, Ajou University
Giltae Song, Pusan National University
Sungroh Yoon, Seoul National University
Jaegul Choo, Korea University
Sael Lee, SUNY Korea / Stony Brook University
Seon Joo Kim, Yonsei University
Hyun Oh Song, Seoul National University

Organizing Committee

Sun Kim (Chair, Seoul National University)
Hyeran Byun (Yonsei University)
Heejoon Chae (Sookmyung Women's University)
Seungjin Choi (Pohang University of Science and Technology)
Geun-Sik Jo (Inha University)
Sung-Bae Jo (Yonsei University)
Daijin Kim (Pohang University of Science and Technology)
Hyunjung Shin (Ajou University)
Sungroh Yoon (Seoul National University

Session 1 Large-Scale Machine Learning

Chair: Alice Oh (KAIST)

#1.1 A Study on Trust Region Update Rules in Newton Methods for Large-scale Linear Classification

Chih-Yang Hsia, National Taiwan University
Ya Zhu, New York University
Chih-Jen Lin, National Taiwan University

The main task in training a linear classifier is to solve an unconstrained minimization problem. In applying an optimization method to obtain a model, typically we iteratively find a good direction and then decide a suitable step size. Past developments of extending optimization methods for large-scale linear classification focus on finding the direction, but little attention has been paid on adjusting the step size. In this work, we explain that inappropriate step-size adjustment may lead to serious slow convergence. Among the two major methods for step-size selection, line search and trust region, we focus on investigating the trust region methods. After presenting some detailed analysis, we develop novel and effective techniques to adjust the trust-region size. Experiments indicate that our new settings significantly outperform existing implementations for large-scale linear classification.

#1.2 Mini-batch Block-coordinate based Stochastic Average Adjusted Gradient Methods to Solve Big Data Problems

Vinod Chauhan, Panjab University Chandigarh
Kalpana Dahiya, Panjab University Chandigarh
Anuj Sharma, Panjab University Chandigarh

Big Data problems in Machine Learning have large number of data points or large number of features, or both, which make training of models difficult because of high computational complexities of single iteration of learning algorithms. To solve such learning problems, Stochastic Approximation offers an optimization approach to make complexity of each iteration independent of number of data points by taking only one data point or mini-batch of data points during each iteration and thereby helping to solve problems with large number of data points. Similarly, Coordinate Descent offers another optimization approach to make iteration complexity independent of the number of features/coordinates/variables by taking only one feature or block of features, instead of all, during an iteration and thereby helping to solve problems with large number of features. In this paper, an optimization framework, namely, Batch Block Optimization Framework has been developed to solve big data problems using the best of Stochastic Approximation as well as the best of Coordinate Descent approaches, independent of any solver. This framework is used to solve strongly convex and smooth empirical risk minimization problem with gradient descent (as a solver) and two novel Stochastic Average Adjusted Gradient methods have been proposed to reduce variance in mini-batch and block-coordinate setting of the developed framework. Theoretical analysis prove linear convergence of the proposed methods and empirical results with bench marked datasets prove the superiority of proposed methods against existing methods.

#1.3 Select-and-Evaluate: A Learning Framework for Large-Scale Knowledge Graph Search

F A Rezaur Rahman Chowdhury, Washington State University
Chao Ma, Oregon State University
Md Rakibul Islam, Washington State University
Mohammad Hossein Namaki, Washington State University
Mohammad Omar Faruk, Washington State University
Janardhan Rao Doppa, Washington State University

Querying graph structured data is a fundamental operation that enables important applications including knowledge graph search, social network analysis, and cyber-network security. However, the growing size of real-world data graphs poses severe challenges for graph search to meet the response-time requirements of the applications. To address these scalability challenges, we develop a learning framework for graph search called {\bf S}ele{\bf c}t-and-Ev{\bf al}uat{\bf e} (SCALE). The key insight is to select a small part of the data graph that is sufficient to answer a given query in order to satisfy the specified constraints on time or accuracy. We formulate the problem of generating the candidate subgraph as a computational search process and induce search control knowledge from training queries using imitation learning. First, we define a search space over candidate selection plans, and identify target selection plans corresponding to the training queries by performing an expensive search. Subsequently, we learn greedy search control knowledge to imitate the search behavior of the target selection plans. Our experiments on large-scale knowledge graphs including DBPedia, YAGO, and Freebase show that using the selection plans generated by the learned search control knowledge, we can significantly improve the computational-efficiency of graph search to achieve high accuracy.

#1.4 Adaptive Sampling Scheme for Learning in Severely Imbalanced Large Scale Data

Wei Zhang, Adobe
Said Kobeissi, Adobe
Scott Tomko, Adobe
Chris Challis, Adobe

Imbalanced data poses a serious challenge for many machine learning and data mining applications. It may significantly affect the performance of learning algorithms. In digital marketing applications, events of interest (positive instances for building predictive models) such as click and purchase are rare. A retail website can easily receive a million visits every day, yet only a small percentage of visits lead to purchase. The large amount of raw data and the small percentage of positive instances make it challenging to build decent predictive models in a timely fashion. In this paper, we propose an adaptive sampling strategy to deal with this problem. It efficiently returns high quality training data, ensures system responsiveness and improves predictive performances.

#1.5 Using Deep Neural Networks to Automate Large Scale Statistical Analysis for Big Data Applications

Rongrong Zhang, Purdue University
Wei Deng, Purdue University
Michael Yu Zhu, Purdue University / Tsinghua University

Statistical analysis (SA) is a complex process to deduce population properties from analysis of data. It usually takes a well-trained analyst to successfully perform SA, and it becomes extremely challenging to apply SA to big data applications. We propose to use deep neural networks to automate the SA process. In particular, we propose to construct convolutional neural networks (CNNs) to perform automatic model

selection and parameter estimation, two most important SA tasks. We refer to the resulting CNNs as the neural model selector and the neural model estimator, respectively, which can be properly trained using labeled data systematically generated from candidate models. Simulation study shows that both the selector and estimator demonstrate excellent performances. The idea and proposed framework can be further extended to automate the entire SA process and have the potential to revolutionize how SA is performed in big data analytics.

#1.6 Accumulated Gradient Normalization

Joeri R. Hermans, Liège University
Gerasimos Spanakis, Maastricht University
Rico Möckel, Maastricht University

This work addresses the instability in asynchronous data parallel optimization. It does so by introducing a novel distributed optimizer with state-of-the-art performance which is able to efficiently optimize a central model under communication constraints. The optimizer achieves this by pushing a normalized sequence of first-order gradients to a parameter server. This implies that the magnitude of the worker deltas are smaller, and provide a better direction towards a minimum compared to first-order gradients. As a result, our approach mitigates the parameter staleness problem more effectively.

#1.7 Efficient Preconditioning for Noisy Separable NMFs by Successive Projection Based Low-Rank Approximations

Tomohiko Mizutani, Tokyo Institute of Technology
Mirai Tanaka, The Institute of Statistical Mathematics

The successive projection algorithm (SPA) can quickly solve a nonnegative matrix factorization problem under a separability assumption. Even if noise is added to the problem, SPA is robust as long as the perturbations caused by the noise are small. In particular, robustness against noise should be high when handling the problems arising from real applications. The preconditioner proposed by Gillis and Vavasis (2015) makes it possible to enhance the noise robustness of SPA. Meanwhile, an additional computational cost is required. The construction of the preconditioner contains a step to compute the top-k truncated singular value decomposition of an input matrix. It is known that the decomposition provides the best rank-k approximation to the input matrix; in other words, a matrix with the smallest approximation error among all matrices of rank less than k. This step is an obstacle to an efficient implementation of the preconditioned SPA. To address the cost issue, we propose a modification of the algorithm for constructing the preconditioner. Although the original algorithm uses the best rank-k approximation, instead of it, our modification uses an alternative. Ideally, this alternative should have high approximation accuracy and low computational cost. To ensure this, our modification employs a rank-k approximation produced by an SPA based algorithm. We analyze the accuracy of the approximation and evaluate the computational cost of the algorithm. We then present an empirical study revealing the actual performance of the SPA.

Session 2 Statistical/Bayesian Machine Learning

Chair: Hyunjung Helen Shin (Ajou University)

#2.1 Probability Calibration Trees

Tim Leathart, University of Waikato
Eibe Frank, University of Waikato
Geoffrey Holmes, University of Waikato
Bernhard Pfahringer, University of Auckland

Obtaining accurate and well calibrated probability estimates from classifiers is useful in many applications, for example, when minimising the expected cost of classifications. Existing methods of calibrating probability estimates are applied globally, ignoring the potential for improvements by applying a more fine-grained model. We propose probability calibration trees, a modification of logistic model trees that can identify regions of the input space in which different probability calibration models should be learned. We compare probability calibration trees to two widely used calibration methods -- isotonic regression and Platt scaling -- and show that our method results in lower root mean squared error on average than both methods, for estimates produced by a variety of base learners.

#2.2 Data sparse nonparametric regression with epsilon-insensitive losses

Maxime Sangnier, UPMC
Olivier Fercoq, Télécom ParisTech
Florence d'Alché Buc, Télécom ParisTech

Leveraging the celebrated support vector regression (SVR) method, we propose a unifying framework in order to deliver regression machines in reproducing kernel Hilbert spaces (RKHSs) with data sparsity. The central point is a new definition of epsilon-insensitivity, valid for many regression losses (including quantile and expectile regression) and their multivariate extensions. We show that the dual optimization problem to empirical risk minimization with epsilon-insensitivity involves a data sparse regularization. We also provide an analysis of the excess of risk as well as a randomized coordinate descent algorithm for solving the dual. Numerical experiments validate our approach.

#2.3 Whitening-Free Least-Squares Non-Gaussian Component Analysis

Hiroaki Shiino, Yahoo Japan Corporation
Hiroaki Sasaki, Nara Institute of Science and Technology
Gang Niu, The University of Tokyo/ RIKEN
Masashi Sugiyama, RIKEN/ The University of Tokyo

Non-Gaussian component analysis (NGCA) is an unsupervised linear dimension reduction method that extracts low-dimensional non-Gaussian "signals" from high-dimensional data contaminated with Gaussian noise. NGCA can be regarded as a generalization of projection pursuit (PP) and independent component analysis (ICA) to multi-dimensional and dependent non-Gaussian components. Indeed, seminal approaches to NGCA are based on PP and ICA. Recently, a novel NGCA approach called least-squares NGCA (LSNGCA) has been developed, which gives a solution analytically through least-squares estimation of log-density gradients and eigendecomposition. However, since pre-whitening of data is involved in LSNGCA, it performs unreliably when the data covariance matrix is ill-conditioned, which is often the case in high-dimensional data

analysis. In this paper, we propose a whitening-free variant of LSNGCA and experimentally demonstrate its superiority.

#2.4 Magnitude-Preserving Ranking for Structured Outputs

Celine Brouard, Aalto university
Eric Bach, Aalto University
Sebastian Böcker, Friedrich-Schiller University
Juho Rousu, Aalto University

In this paper, we present a novel method for solving structured prediction problems, based on combining Input Output Kernel Regression (IOKR) with an extension of magnitude-preserving ranking to structured output spaces. In particular, we concentrate on the case where a set of candidate outputs has been given, and the associated pre-image problem calls for ranking the set of candidate outputs. Our method, called magnitude-preserving IOKR, both aims to produce a good approximation of the output feature vectors, and to preserve the magnitude differences of the output features in the candidate sets. For the case where the candidate set does not contain corresponding 'correct' inputs, we propose a method for approximating the inputs through application of IOKR in the reverse direction. We apply our method to two learning problems: cross-lingual document retrieval and metabolite identification. Experiments show that the proposed approach improves performance over IOKR, and in the latter application obtains the current state-of-the-art accuracy.

#2.5 A Word Embeddings Informed Focused Topic Model

He Zhao, Monash University
Lan Du, Monash University
Wray Buntine, Monash University

In natural language processing and related fields, it has been shown that the word embeddings can successfully capture both the semantic and syntactic features of words. They can serve as complementary information to topics models, especially especially for the cases where word co-occurrence data is insufficient, such as with short texts. In this paper, we propose a focused topic model where how a topic focuses on words is informed by word embeddings. Our models is able to discover more informed focused topics with more representative words, leading to better modelling accuracy and topic quality. With the data argumentation technique, we can derive an efficient Gibbs sampling algorithm, which benefits from the fully local conjugacy of the model. We conduct extensive experiments on several real world datasets demonstrate that our model achieves comparable or improved performance in terms of both perplexity and topic coherence, particularly in handling sparse text data.

#2.6 A Mutually-Dependent Hadamard Kernel for Modelling Latent Variable Couplings

Sami Remes, Aalto University
Markus Heinonen, Aalto University
Samuel Kaski, Aalto University

We introduce a novel kernel that models input-dependent couplings across multiple latent processes. The pairwise kernel measures covariance both along inputs and across different latent signals in a mutually-dependent fashion. The latent correlation Gaussian process (LCGP) model combines these non-stationary latent components into multiple outputs by an input-dependent mixing matrix. Probit classification and support for multiple observation sets are derived by Variational Bayesian inference.

Results on several datasets indicate that the LCGP model can recover the correlations between latent signals while simultaneously achieving state-of-the-art performance. We highlight the latent covariances with an EEG classification dataset where latent brain processes and their couplings simultaneously emerge from the model.

#2.7 Recovering Probability Distributions from Missing Data

Jin Tian, Iowa State University

A probabilistic query may not be estimable from observed data corrupted by missing values if the data are not missing at random (MAR). It is therefore of theoretical interest and practical importance to determine in principle whether a probabilistic query is estimable from missing data or not when the data are not MAR. We present algorithms that systematically determine whether the joint probability distribution or a target marginal distribution is estimable from observed data with missing values, assuming that the data-generation model is represented as a Bayesian network, known as m-graphs, that not only encodes the dependencies among the variables but also explicitly portrays the mechanisms responsible for the missingness process. The results significantly advance the existing work.

Session 3 Machine Learning Applications

Chair: Masashi Sugiyama (RIKEN/The University of Tokyo)

#3.1 PHD: A Probabilistic Model of Hybrid Deep Collaborative Filtering for Recommender Systems

Jie Liu, Shanghai Jiao Tong University
Dong Wang, Shanghai Jiao Tong University
Yue Ding, Shanghai Jiao Tong University

Collaborative Filtering (CF), a well-known approach in producing recommender systems, has achieved wide use and excellent performance not only in research but also in industry. However, problems related to cold start and data sparsity have caused CF to attract an increasing amount of attention in efforts to solve these problems. Traditional approaches adopt side information to extract effective latent factors but still have some room for growth. Due to the strong characteristic of feature extraction in deep learning, many researchers have employed it with CF to extract effective representations and to enhance its performance in rating prediction. Based on this previous work, we propose a probabilistic model that combines a stacked denoising autoencoder and a convolutional neural network together with auxiliary side information (i.e, both from users and items) to extract users and items' latent factors, respectively. Extensive experiments for four datasets demonstrate that our proposed model outperforms other traditional approaches and deep learning models making it state of the art.

#3.2 Recognizing Art Style Automatically in painting with deep learning

Adrian Lecoutre, INSA Rouen
Florian Yger, Université Paris-Dauphine
Benjamin Negrevergne, Université Paris-Dauphine

An art style (or movement) refers to a tendency in an art with a specific goal or philosophy and sometimes setting some explicit canon, followed by a group of artists during a restricted period of time. Hence, art style is probably the most general descriptor that can be used to classify a painting as it merges visual and historical information. In the context of large visual art databases, style is therefore a crucial metadata for the description of paintings. In this article, in an attempt at automatically detecting the art style of a painting from an image, we propose to apply a deep learning approach and illustrate our study with results on WikiArt data.

#3.3 Computer Assisted Composition with Recurrent Neural Networks

Christian Walder, DATA61
Dongwoo Kim, ANU

Sequence modeling with neural networks has lead to powerful models of symbolic music data. We address the problem of exploiting these models to reach creative musical goals. To this end we generalise previous work, which sampled Markovian sequence models under the constraint that the sequence belong to the language of a given finite state machine. We consider more expressive non-Markov models, thereby requiring approximate sampling which we provide in the form of an efficient sequential Monte Carlo method. In addition we provide and compare with a beam search strategy for conditional probability maximisation. Our algorithms are capable of convincingly re-harmonising famous

musical works. To demonstrate this we provide visualisations, quantitative experiments, a human listening test and illustrative audio examples. We find both the sampling and optimisation procedures to be effective, yet complementary in character. For the case of highly permissive constraint sets, we find that sampling is to be preferred due to the overly regular nature of the optimisation based results.

#3.4 Learning Deep Semantic Embeddings for Cross-Modal Retrieval

Cuicui Kang, Institute of Information Engineering, CAS
Shengcai Liao, Institute of Automation, CAS
Zhen Li, Institute of Information Engineering, CAS
Zigang Cao, Institute of Information Engineering, CAS
Gang Xiong, Institute of Information Engineering, CAS

Deep learning methods have been actively researched for cross-modal retrieval, with the softmax cross-entropy loss commonly applied for supervised learning. However, the softmax cross-entropy loss is known to result in large intra-class variances, which is not not very suited for cross-modal matching. In this paper, a deep architecture called Deep Semantic Embedding (DSE) is proposed, which is trained in an end-to-end manner for image-text cross-modal retrieval. With images and texts mapped to a feature embedding space, class labels are used to guide the embedding learning, so that the embedding space has a semantic meaning common for both images and texts. This way, the difference between different modalities is eliminated. Under this framework, the center loss is introduced beyond the commonly used softmax cross-entropy loss to achieve both inter-class separation and intra-class compactness. Besides, a distance based softmax cross-entropy loss is proposed to jointly consider the softmax cross-entropy and center losses in fully gradient based learning. Experiments have been done on three popular image-text cross-modal retrieval databases, showing that the proposed algorithms have achieved the best overall performances.

#3.5 Pyramid Person Matching Network for Person Re-identification

Chaojie Mao, Zhejiang University
Yingming Li, Zhejiang University
Zhongfei Zhang, Zhejiang University
Yaqing Zhang, Zhejiang University
Xi Li, Zhejiang University

In this work, we present a deep convolutional pyramid person matching network (PPMN) with specially designed Pyramid Matching Module to address the problem of person re-identification. The architecture takes a pair of RGB images as input, and outputs a similiarity value indicating whether the two input images represent the same person or not. Based on deep convolutional neural networks, our approach first learns the discriminative semantic representation with the semantic-component-aware features for persons and then employs the Pyramid Matching Module to match the common semantic-components of persons, which is robust to the variation of spatial scales and misalignment of locations posed by viewpoint changes. The above two processes are jointly optimized via a unified end-to-end deep learning scheme. Extensive experiments on several benchmark datasets demonstrate the effectiveness of our approach against the state-of-the-art approaches, especially on the rank-1 recognition rate.

#3.6 Radical-level Ideograph Encoder for RNN-based Sentiment Analysis of Chinese and Japanese

Yuanzhi Ke, Keio University
Masafumi Hagiwara, Keio University

The character vocabulary can be very large in non-alphabetic languages such as Chinese and Japanese, which makes neural network models huge to process such languages. We explored a model for sentiment classification that takes the embeddings of the radicals of the Chinese characters, i.e, hanzi of Chinese and kanji of Japanese. Our model is composed of a CNN word feature encoder and a bi-directional RNN document feature encoder. The results achieved are on par with the character embedding-based models, and close to the state-of-the-art word embedding-based models, with 90% smaller vocabulary, and at least 13% and 80% fewer parameters than the character embedding-based models and word embedding-based models respectively. The results suggest that the radical embedding-based approach is cost-effective for machine learning on Chinese and Japanese.

#3.7 Attentive Path Combination for Knowledge Graph Completion

Xiaotian Jiang, Institute of Information Engineering / School of Cyber Security, CAS
Quan Wang, Institute of Information Engineering / School of Cyber Security / State Key Laboratory of Information Security, CAS
Baoyuan Qi, Institute of Information Engineering / School of Cyber Security, CAS
Yongqin Qiu, Institute of Information Engineering / School of Cyber Security, CAS
Peng Li, Institute of Information Engineering / School of Cyber Security, CAS
Bin Wang, Institute of Information Engineering / School of Cyber Security, CAS

Knowledge Graphs (KGs) are often greatly incomplete, necessitating a demand for KG completion. Path-based relation inference is one of the most important approaches to this task. Traditional method treats each path between entity pairs as an atomic feature, thus inducing sparsity. Recently, neural network models solved this problem by decomposing path as a sequence of all relations in the path, so as to model the path representation using Recurrent Neural Network (RNN) architectures. In cases that there are multiple paths between an entity pair, state-of-the-art neural models either select just one path with the largest score, or make usage of simple score pooling methods like Top-K, Average, LogSumExp. Unfortunately, none of these methods is good enough to support relation inference that only bases on collecting all the evidences from multiple informative paths. In this paper, we propose a novel path-based relation inference model that learns effective entity pair representation with attentive path combination method. Given an entity pair and a set of paths connecting the pair in KG, our model allows for integrating information from each of the paths, and form a dynamic entity pair representation with respect to each candidate relation to query. We empirically evaluate the proposed method on a real-world dataset. Experimental results show that the proposed model achieves better performance than state-of-the-art path-based relation inference methods.

Session 4 Weakly-supervised/Unsupervised Learning

Chair: Paul Weng (SYSU-CMU)

#4.1 ST-GAN: Unsupervised Image Semantic Transformation Using Generative Adversarial Networks

JiChao Zhang, Shandong University
Fan Zhong, Shandong University
Gongze Cao, Zhejiang University
Xueying Qin, Shandong University

Image semantic transformation aims to convert one image into another image with different semantic features (e.g., face pose, hairstyle). The previous methods, which learn the mapping function from one image domain to the other, require supervised information directly or indirectly. In this paper, we propose an unsupervised image semantic transformation method called semantic transformation generative adversarial networks (ST-GAN). We further improve ST-GAN with the Wasserstein distance to generate more realistic images and propose a method called local mutual information maximization to obtain a more explicit semantic transformation. ST-GAN has the ability to map the image semantic features into the latent vector and then perform transformation by controlling the latent vector. After the proposed framework is trained on a benchmark, the original face images can be reconstructed and then translated into various images with different semantic features.

#4.2 One Class Splitting Criteria for Random Forests

Nicolas Goix, Télécom ParisTech
Nicolas Drougard, ISAE
Romain Brault, Télécom ParisTech
Maël Chiapino, Télécom ParisTech

Random Forests (RFs) are strong machine learning tools for classification and regression. However, they remain supervised algorithms, and no extension of RFs to the one-class setting has been proposed, except for techniques based on second-class sampling. This work fills this gap by proposing a natural methodology to extend standard splitting criteria to the one-class setting, structurally generalizing RFs to one-class classification. An extensive benchmark of seven state-of-the-art anomaly detection algorithms is also presented. This empirically demonstrates the relevance of our approach.

#4.3 Semi-supervised Convolutional Neural Networks for Identifying Wi-Fi Interference Sources

Krista Longi, University of Helsinki
Teemu Pulkkinen, University of Helsinki
Arto Klami, University of Helsinki

We present a convolutional neural network for identifying radio frequency devices from signal data, in order to detect possible interference sources for wireless local area networks. Collecting training data for this problem is particularly challenging due to a high number of possible interfering devices, difficulty in obtaining precise timings, and the need to measure the devices in varying conditions. To overcome this challenge we focus on semi-supervised learning, aiming to minimize the need to of reliable training samples while utilizing larger amounts of unsupervised labels to improve the accuracy. In particular, we propose a novel structured extension

of the pseudo-label technique to take advantage of temporal continuity in the data and show that already a few seconds of training data for each device is sufficient for highly accurate recognition.

#4.4 Learning RBM with a DC programming Approach

Vidyadhar Upadhya, Indian Institute of Science
BP S Sastry, Indian Institute of Science

By exploiting the property that the RBM log-likelihood function is the difference of convex functions, we formulate a stochastic variant of the difference of convex functions (DC) programming to minimize the negative log-likelihood. Interestingly, the traditional contrastive divergence algorithm is a special case of the above formulation and the hyperparameters of the two algorithms can be chosen such that the amount of computation per mini-batch is identical. We show that for a given computational budget the proposed algorithm almost always reaches a higher log-likelihood more rapidly, compared to the standard contrastive divergence algorithm. Further, we modify this algorithm to use the centered gradients and show that it is more efficient and effective compared to the standard centered gradient algorithm on benchmark datasets.

#4.5 Learning Safe Multi-Label Prediction for Weakly Labeled Data

Tong Wei, Nanjing University
Lan-Zhe Guo, Nanjing University
Yu-Feng Li, Nanjing University
Wei Gao, Nanjing University

In this paper we study multi-label learning with weakly labeled data, i.e., labels of training examples are incomplete, which commonly occurs in real applications, e.g., image classification, document categorization. This setting includes, e.g., (i) \emph{semi-supervised multi-label learning} where completely labeled examples are partially known
(ii) \emph{weak label learning} where relevant labels of examples are partially known
iii) \emph{extended weak label learning} where relevant and irrelevant labels of examples are partially known. Previous studies often expect that the learning method with the use of weakly labeled data will improve the performance, as more data are employed. This, however, is not always the cases in reality, i.e., weakly labeled data may sometimes degenerate the learning performance. It is desirable to learn \emph{safe} multi-label prediction that will not hurt performance when weakly labelled data is involved in the learning procedure. In this work we optimize multi-label evaluation metrics (F$_1$ score and Top-k precision) given that the ground-truth label assignment is realized by a convex combination of base multi-label learners. To cope with the infinite number of possible ground-truth label assignments, cutting-plane strategy is adopted to iteratively generate the most helpful label assignments. The whole optimization is cast as a series of simple linear programs in an efficient manner. Extensive experiments on three weakly labeled learning tasks, namely, i) semisupervised multi-label learning
ii) weak-label learning and iii) extended weak-label learning, clearly show that our proposal improves the safeness of using weakly labelled data compared with many state-of-the-art methods.

#4.6 Semi-Supervised AUC Optimization based on Positive-Unlabeled Learning

Tomoya Sakai, The University of Tokyo/ RIKEN
Gang Niu, The University of Tokyo/ RIKEN
Masashi Sugiyama, RIKEN / The University of Tokyo

Maximizing the area under the receiver operating characteristic curve (AUC) is a standard approach to imbalanced classification. So far, various supervised AUC optimization methods have been developed and they are also extended to semisupervised scenarios to cope with small sample problems. However, existing semisupervised AUC optimization methods rely on strong distributional assumptions, which are rarely satisfied in real-world problems. In this paper, we propose a novel semisupervised AUC optimization method that does not require such restrictive assumptions. We first develop an AUC optimization method based only on positive and unlabeled data (PU-AUC) and then extend it to semi-supervised learning by combining it with a supervised AUC optimization method. We theoretically prove that, without the restrictive distributional assumptions, unlabeled data contribute to improving the generalization performance in PU and semi-supervised AUC optimization methods. Finally, we demonstrate the practical usefulness of the proposed methods through experiments.

#1.1: A Study on Trust Region Update Rules in Newton Methods for Large-scale Linear Classification

#1.2: Mini-batch Block-coordinate based Stochastic Average Adjusted Gradient Methods to Solve Big Data Problems

#1.3: Select-and-Evaluate: A Learning Framework for Large-Scale Knowledge Graph Search

#1.4: Adaptive Sampling Scheme for Learning in Severely Imbalanced Large Scale Data

#1.5: Using Deep Neural Networks to Automate Large Scale Statistical Analysis for Big Data Applications

#1.6: Accumulated Gradient Normalization

#1.7: Efficient Preconditioning for Noisy Separable NMFs by Successive Projection Based Low-Rank Approximations

#2.1: Probability Calibration Trees

#2.2: Data Sparse Nonparametric Regression with Epsilon-Insensitive Losses

#2.3: Whitening-Free Least-Squares Non-Gaussian Component Analysis

#2.4: Magnitude-Preserving Ranking for Structured Outputs

#2.5: A Word Embeddings Informed Focused Topic Model

#2.6: A Mutually-Dependent Hadamard Kernel for Modelling Latent Variable Couplings

#2.7: Recovering Probability Distributions from Missing Data

#3.1: PHD: A Probabilistic Model of Hybrid Deep Collaborative Filtering for Recommender Systems

#3.2: Recognizing Art Style Automatically in Painting with Deep Learning

#3.3: Computer Assisted Composition with Recurrent Neural Networks

#3.4: Learning Deep Semantic Embeddings for Cross-Modal Retrieval

#3.5: Pyramid Person Matching Network for Person Re-identification

#3.6: Radical-level Ideograph Encoder for RNN-based Sentiment Analysis of Chinese and Japanese

#3.7: Attentive Path Combination for Knowledge Graph Completion

#4.1: ST-GAN: Unsupervised Facial Image Semantic Transformation Using Generative Adversarial Networks

#4.2: One Class Splitting Criteria for Random Forests

#4.3: Semi-supervised Convolutional Neural Networks for Identifying Wi-Fi Interference Sources

#4.4: Learning RBM with a DC programming Approach

#4.5: Learning Safe Multi-Label Prediction for Weakly Labeled Data

#4.6: Semi-Supervised AUC Optimization based on Positive-Unlabeled Learning

Session 5 Deep Learning

Chair: Jihun Hamm (Ohio State University)

#5.1 Learning Convolutional Neural Networks using Hybrid Orthogonal Projection and Estimation

Hengyue Pan, York University
Hui Jiang, York University

Convolutional neural networks (CNNs) have yielded the excellent performance in a variety of computer vision tasks, where CNNs typically adopt a similar structure consisting of convolution layers, pooling layers and fully connected layers. In this paper, we propose to apply a novel method, namely Hybrid Orthogonal Projection and Estimation (HOPE), to CNNs in order to introduce orthogonality into the CNN structure. The HOPE model can be viewed as a hybrid model to combine feature extraction using orthogonal linear projection with mixture models. It is an effective model to extract useful information from the original high-dimension feature vectors and meanwhile filter out irrelevant noises. In this work, we present three different ways to apply the HOPE models to CNNs, i.e., {\em HOPE-Input}, {\em single-HOPE-Block} and {\em multi-HOPE-Blocks}. For {\em HOPE-Input} CNNs, a HOPE layer is directly used right after the input to de-correlate high-dimension input feature vectors. Alternatively, in {\em single-HOPE-Block} and {\em multi-HOPE-Blocks} CNNs, we consider to use HOPE layers to replace one or more blocks in the CNNs, where one block may include several convolutional layers and one pooling layer. The experimental results on CIFAR-10, CIFAR-100 and ImageNet databases have shown that the orthogonal constraints imposed by the HOPE layers can significantly improve the performance of CNNs in these image classification tasks (we have achieved one of the best performance when image augmentation has not been applied, and top 5 performance with image augmentation).

#5.2 Limits of End-to-End Learning

Tobias Glasmachers, Ruhr-University Bochum

End-to-end learning refers to training a possibly complex learning system by applying gradient-based learning to the system as a whole. End-to-end learning systems are specifically designed so that all modules are differentiable. In effect, not only a central learning machine, but also all "peripheral" modules like representation learning and memory formation are covered by a holistic learning process. The power of end-to-end learning has been demonstrated on many tasks, like playing a whole array of Atari video games with a single architecture. While pushing for solutions to more challenging tasks, network architectures keep growing more and more complex. In this paper we ask the question whether and to what extent end-to-end learning is a future-proof technique in the sense of scaling to complex and diverse data processing architectures. We point out potential inefficiencies, and we argue in particular that end-to-end learning does not make optimal use of the modular design of present neural networks. Our surprisingly simple experiments demonstrate these inefficiencies, up to the complete breakdown of learning.

#5.3 Locally Smoothed Neural Networks

Liang Pang, Institute of Computing Technology, CAS
Yanyan Lan, Institute of Computing Technology, CAS
Jun Xu, Institute of Computing Technology, CAS
Jiafeng Guo, Institute of Computing Technology, CAS
Xueqi Cheng, Institute of Computing Technology, CAS

Convolutional Neural Networks (CNN) and the locally connected layer are limited in capturing the importance and relations of different local receptive fields, which are often crucial for tasks such as face verification, visual question answering, and word sequence prediction. To tackle the issue, we propose a novel locally smoothed neural network (LSNN) in this paper. The main idea is to represent the weight matrix of the locally connected layer as the product of the kernel and the smoother, where the kernel is shared over different local receptive fields, and the smoother is for determining the importance and relations of different local receptive fields. Specifically, a multi-variate Gaussian function is utilized to generate the smoother, for modeling the location relations among different local receptive fields. Furthermore, the content information can also be leveraged by setting the mean and precision of the Gaussian function according to the content. Experiments on some variant of MNIST clearly show our advantages over CNN and locally connected layer.

#5.4 Deep Competitive Pathway Networks

Jia-Ren Chang, National Chiao Tung University
Yong-Sheng Chen, National Chiao Tung University

In the design of deep neural architectures, recent studies have demonstrated the benefits of grouping subnetworks into a larger network. For examples, the Inception architecture integrates multi-scale subnetworks and the residual network can be regarded that a residual unit combines a residual subnetwork with an identity shortcut. In this work, we embrace this observation and propose the Competitive Pathway Network (CoPaNet). The CoPaNet comprises a stack of competitive pathway units and each unit contains multiple parallel residual-type subnetworks followed by a max operation for feature competition. This mechanism enhances the model capability by learning a variety of features in subnetworks. The proposed strategy explicitly shows that the features propagate through pathways in various routing patterns, which is referred to as pathway encoding of category information. Moreover, the cross-block shortcut can be added to the CoPaNet to encourage feature reuse. We evaluated the proposed CoPaNet on four object recognition benchmarks: CIFAR-10, CIFAR-100, SVHN, and ImageNet. CoPaNet obtained the state-of-the-art or comparable results using similar amounts of parameters.

#5.5 Scale-Invariant Recognition by Weight-Shared CNNs in Parallel

Ryo Takahashi, Kobe University
Takashi Matsubara, Kobe University
Kuniaki Uehara, Kobe University

Deep convolutional neural networks (CNNs) have become one of the most successful methods for image processing tasks in past few years. Recent studies on modern residual architectures, enabling CNNs to be much deeper, have achieved much better results thanks to their high expressive ability by numerous parameters. In general, CNNs are known to have the robustness to the small parallel shift of objects in images by their local receptive fields, weight parameters shared by each unit, and pooling layers sandwiching them. However, CNNs have a limited robustness to the other geometric transformations such as scaling and rotation, and this lack becomes an obstacle to performance

improvement even now. This paper proposes a novel network architecture, the weight-shared multi-stage network (WSMS-Net), and focuses on acquiring the scale invariance by constructing of multiple stages of CNNs. The WSMS-Net is easily combined with existing deep CNNs, enables existing deep CNNs to acquire a robustness to the scaling, and therefore, achieves higher classification accuracy on CIFAR-10, CIFAR-100 and ImageNet datasets.

#5.6 Nested LSTMs

Joel Ruben Antony Moniz, Carnegie Mellon University
David Krueger, Université de Montreal

We propose Nested LSTMs (NLSTM), a novel RNN architecture with multiple levels of memory. Nested LSTMs add depth to LSTMs via nesting as opposed to stacking. The value of a memory cell in an NLSTM is computed by an LSTM cell, which has its own inner memory cell. Specifically, instead of computing the value of the (outer) memory cell as $c^{outer}_t = f_t \odot c_{t-1} + i_t \odot g_t$, NLSTM memory cells use the concatenation $(f_t \odot c_{t-1}, i_t \odot g_t)$ as input to an inner LSTM (or NLSTM) memory cell, and set c^{outer}_t = h^{inner}_t. Nested LSTMs outperform both stacked and single-layer LSTMs with similar numbers of parameters in our experiments on various character-level language modeling tasks, and the inner memories of an LSTM learn longer term dependencies compared with the higher-level units of a stacked LSTM.

#5.7 Neural-Power: Predict and Deploy Energy-Efficient Convolutional Neural Networks

Ermao Cai, Carnegie Mellon University
Da-Cheng Juan, Google Research
Dimitrios Stamoulis, Carnegie Mellon University
Diana Marculescu, Carnegie Mellon University

"How much energy is consumed for an inference made by a convolutional neural network (CNN)?" With the increased popularity of CNN models deployed on the wide-spectrum of platforms (from mobile devices to workstations), the energy consumption of making an inference with a CNN model has drawn significantly more attention. From lengthening battery life of mobile devices to reducing the energy bill of a datacenter, it is important to understand the energy efficiency of CNN models during serving for making an inference, without actually training the model. In this work, we propose Neural-Power: a layer-wise predictive framework based on sparse polynomial regression, for predicting the serving energy consumption of a CNN model deployed on any GPU platform. Given the architecture (or hyper-parameters) of a CNN model, the proposed framework provides an accurate prediction and breakdown for both power and runtime across all layers in the whole network, aiming at helping machine learners quickly identify the bottleneck(s) of execution time or power consumption. We also propose a metric "energy-precision ratio" (EPR) to guide machine learners in selecting an energy-efficient CNN architecture that better trades off the energy consumption and prediction accuracy. The experimental results show that the prediction accuracy of the proposed Neural-Power outperforms the best published model to date, yielding an improvement in accuracy up to 68.5%. We also assess the accuracy of predictions at the network level, by predicting the runtime, power, and energy of state-of-the-art CNN configurations, achieving an average accuracy of 88.24% in runtime, 88.34% in power, and 97.21% in energy. We comprehensively corroborate the effectiveness of Neural-Power as a powerful framework for machine learning practitioners by testing it on different GPU platforms and Deep Learning software tools.

Session 6 Learning Theory

Chair: Kee-Eung Kim (KAIST)

#6.1 Regret For Expected Improvement Over The Best-Observed Value and Stopping Condition

Vu Nguyen, Deakin University
Sunil Gupta, Deakin University
Santu Rana, Deakin University
Cheng Li, Deakin University
Svetha Venkatesh, Deakin University

Bayesian optimization (BO) is a sample-efficient method for global optimization of expensive, noisy, black-box functions using probabilistic methods. The performance of a Bayesian optimization method depends on its selection strategy through the acquisition function. Expected improvement (EI) is one of the most widely used acquisition functions for BO that finds the expectation of the improvement function over the incumbent. The incumbent is usually selected as the best-observed value so far, termed as y^max (for the maximizing problem). Recent work has studied the convergence rate for EI under some mild assumptions or zero noise of observations. Especially, the work of (Wang and de Freitas, 2014) has derived the sublinear regret for EI under a stochastic noise. However, due to the difficulty in stochastic noise setting and to make the convergent proof feasible, they use an alternative choice for the incumbent as the maximum of the Gaussian process predictive mean, μ^max . This modification makes the algorithm computationally inefficient because it requires an additional global optimization step to estimate μ^max that is costly and may be inaccurate. To address this issue, we derive a sublinear convergence rate for EI using the commonly used y^max . Moreover, our analysis is the first to study a stopping criteria for EI to prevent unnecessary evaluations. Our analysis complements the results of (Wang and de Freitas, 2014) to theoretically cover two incumbent settings for EI. Finally, we demonstrate empirically that EI using y max is both more computationally efficiency and more accurate than EI using μ^max .

#6.2 Distributionally Robust Groupwise Regularization Estimator

Jose Blanchet, Columbia University
Yang Kang, Columbia University

Regularized estimators in the context of group variables have been applied successfully in model and feature selection in order to preserve interpretability. We formulate a Distributionally Robust Optimization (DRO) problem which recovers popular estimators, such as Group Square Root Lasso (GSRL). Our DRO formulation allows us to interpret GSRL as a game, in which we learn a regression parameter while an adversary chooses a perturbation of the data. We wish to pick the parameter to minimize the expected loss under any plausible model chosen by the adversary - who, on the other hand, wishes to increase the expected loss. The regularization parameter turns out to be precisely determined by the amount of perturbation on the training data allowed by the adversary. In this paper, we introduce a data-driven (statistical) criterion for the optimal choice of regularization, which we evaluate asymptotically, in closed form, as the size of the training set increases. Our easy-to-evaluate regularization formula is compared against cross-validation, showing good (sometimes superior) performance.

#6.3 Rate Optimal Estimation for High Dimensional Spatial Covariance Matrices

Aidong Adam Ding, Northeastern University
Yi Li, Northeastern University
Jennifer Dy, Northeastern University

Spatial covariance matrix estimation is of great significance in many applications in climatology, econometrics and many other fields with complex data structures involving spatial dependencies. High dimensionality brings new challenges to this problem, and no theoretical optimal estimator has been proved for the spatial high-dimensional covariance matrix. Over the past decade, the method of regularization has been introduced to high-dimensional covariance estimation for various structured matrices, to achieve rate optimal estimators. In this paper, we aim to bridge the gap in these two research areas. We use a structure of block bandable covariance matrices to incorporate spatial dependence information, and study rate optimal estimation of this type of structured high dimensional covariance matrices. A double tapering estimator is proposed, and is shown to achieve the asymptotic minimax error bound. Numerical studies on both synthetic and real data are conducted showing the improvement of the double tapering estimator over the sample covariance matrix estimator.

#6.4 A Quantum-Inspired Ensemble Method and Quantum-Inspired Forest Regressors

Zeke Xie, The University of Tokyo
Issei Sato, The University of Tokyo

We propose a Quantum-Inspired Subspace(QIS) Ensemble Method for generating feature ensembles based on feature selections. We assign each principal component a Fraction Transition Probability as its probability weight based on Principal Component Analysis and quantum interpretations. In order to generate the feature subset for each base regressor, we select a feature subset from principal components based on Fraction Transition Probabilities. The idea originating from quantum mechanics can encourage ensemble diversity and the accuracy simultaneously. We incorporate Quantum-Inspired Subspace Method into Random Forest and propose Quantum-Inspired Forest. We theoretically prove that the quantum interpretation corresponds to the first order approximation of ensemble regression. We also evaluate the empirical performance of Quantum-Inspired Forest and Random Forest in multiple hyperparameter settings. Quantum-Inspired Forest prove the significant robustness of the default hyperparameters on most data sets. The contribution of this work is two-fold, a novel ensemble regression algorithm inspired by quantum mechanics and the theoretical connection between quantum interpretations and machine learning algorithms.

#6.5 On the Flatness of Loss Surface for Two-layered ReLU Networks

Jiezhang Cao, South China University of Technology
Qingyao Wu, South China University of Technology
Yuguang Yan, South China University of Technology
Li Wang, University of Texas at Arlington
Mingkui Tan, South China University of Technology

Deep learning has achieved unprecedented practical success in many applications. Despite its empirical success, however, the theoretical understanding of deep neural networks still remains a major open problem. In this paper, we explore properties of two-layered ReLU network. For simplicity, we assume that the optimal model parameters (also called ground-truth parameters) are known. We then assume that a network receives Gaussian input and is trained by minimizing the expected squared loss

between learned parameters and known ground-truth parameters. Theoretically, we propose a normal equation for critical points, and study the invariances under three kinds of transformations. We prove that these transformations can keep the loss of a critical point invariant, thus can form flat regions. Therefore, how to escape from flat regions is vital for training networks.

#6.6 A Covariance Matrix Adaptation Evolution Strategy for Direct Policy Search in Reproducing Kernel Hilbert Space

Ngo Anh Vien, Queen's University of Belfast
Viet-Hung Dang, Duy Tan University
Chung TaeChoong, Kyung Hee University

The covariance matrix adaptation evolution strategy (CMA-ES) is an efficient derivative-free optimization algorithm. It optimizes a black-box objective function over a well defined parameter space. In some problems, such parameter spaces are defined using function approximation in which feature functions are manually defined. Therefore, the performance of those techniques strongly depends on the quality of chosen features. Hence, enabling CMA-ES to optimize on a more complex and general function class of the objective has long been desired. Specifically, we consider modeling the input space for black-box optimization in reproducing kernel Hilbert spaces (RKHS). This modeling leads to a functional optimization problem whose domain is a function space that enables us to optimize in a very rich function class. In addition, we propose CMA-ES-RKHS, a generalized CMA-ES framework, that performs black-box functional optimization in the RKHS. A search distribution, represented as a Gaussian process, is adapted by updating both its mean function and covariance operator. Adaptive representation of the function and covariance operator is achieved with sparsification techniques. We evaluate CMA-ES-RKHS on a simple functional optimization problem and benchmark reinforcement learning (RL) domains. For an application in RL, we model policies for MDPs in RKHS and transform a cumulative return objective as a functional of RKHS policies, which can be optimized via CMA-ES-RKHS. This formulation results in a black-box functional policy search framework.

Session 7 Multi-view, Multi-task, and Crowdsouced Learning

Chair: Yu-Feng Li (Nanjing University)

#7.1 Instance Specific Discriminative Modal Pursuit: A Serialized Approach

Yang Yang, Nanjing University
De-Chuan Zhan, Nanjing University
Ying Fan, Nanjing University
Yuan Jiang, Nanjing University

With the fast development of data collection techniques, a huge amount of complex multi-modal data are generated, shared and stored on the Internet. The burden of extracting multi-modal features for test instances in data analysis becomes the main fact that hurts the efficiency of prediction. In this paper, in order to reduce the modal extraction cost in serialized classification system, we propose a novel end-to-end serialized adaptive decision approach named Discriminative Modal Pursuit (DMP), which can automatically extract instance-specifically discriminative modal sequence for reducing the cost of feature extraction in the test phase. Rather than jointly optimize a highly non-convex empirical risk minimization problem, we are inspired by LSTM, and the proposed DMP can turn to learn the decision policies which predict the label information and decide the modalities to be extracted simultaneously within limited modal acquisition budget. Consequently, DMP approach can balance the classification performance and modal feature extraction cost by utilizing different modalities for different test instances. Empirical studies show that DMP is more efficient and effective than existing modal/feature extraction methods.

#7.2 Multi-view Clustering with Adaptively Learned Graph

Hong Tao, National University of Defense Technology
Chenping Hou, National University of Defense Technology
Jubo Zhu, National University of Defense Technology
Dongyun Yi, National University of Defense Technology

Multi-view clustering, which aims to improve the clustering performance by exploring the data's multiple representations, has become an importance research direction. Graph based methods have been widely studied and achieve promising performance for multi-view clustering. However, most existing multi-view graph based methods perform clustering on the fixed input graphs, and the results are dependent on the quality of input graphs. In this paper, instead of fixing the input graphs, we propose Multi-view clustering with Adaptively Learned Graph (MALG), learning a new common similarity matrix. In our model, we not only consider the importance of multiple graphs from view level, but also focus on the performance of similarities within a view from sample-pair level. Sample-pair-specific weights are introduced to exploit the connection across views in more depth. In addition, the obtained optimal graph can be partitioned into specific clusters directly, according to its connected components. Experimental results on toy and real-world datasets demonstrate the efficacy of the proposed algorithm.

#7.3 Learning Predictive Leading Indicators for Forecasting Time Series Systems with Unknown Clusters of Forecast Tasks

Magda Gregorová, University of Applied Sciences of Western Switzerland / University of Geneva
Alexandros Kalousis, University of Applied Sciences of Western Switzerland / University of Geneva
Stéphan Marchand-Maillet, University of Geneva

We present a new method for forecasting systems of multiple interrelated time series. The method learns the forecast models together with discovering leading indicators from within the system that serve as good predictors improving the forecast accuracy and a cluster structure of the predictive tasks around these. The method is based on the classical linear vector autoregressive model (VAR) and links the discovery of the leading indicators to inferring sparse graphs of Granger causality. We formulate a new constrained optimisation problem to promote the desired sparse structures across the models and the sharing of information amongst the learning tasks in a multi-task manner. We propose an algorithm for solving the problem and document on a battery of synthetic and real-data experiments the advantages of our new method over baseline VAR models as well as the state-of-the-art sparse VAR learning methods.

#7.4 Multi-Task Structured Prediction for Entity Analysis: Search-Based Learning Algorithms

Chao Ma, Oregon State University
Janardhan Rao Doppa, Washington State University
Prasad Tadepalli, Oregon State University
Hamed Shahbazi, Oregon State University
Xiaoli Fern, Oregon State University

We study several search-based approaches to the problem of multi-task structured prediction (MTSP) in the context of multiple entity analysis tasks in natural language processing. In search-based structured prediction, we learn a scoring function from supervised training data where the best solutions correspond to the highest scores. We study 3 different search architectures to multi-task structured prediction that make different tradeoffs between speed and accuracy. In the fastest "pipeline" architecture, we solve different tasks one after another in a pipelined fashion. In the "joint" architecture, which is the slowest, we formulate MTSP as a single-task structured prediction, and search the joint search space. We introduce two different ways of pruning the search space to make search more efficient. In the intermediate "cyclic" architecture, we cycle through the tasks in sequence multiple times until there is no performance improvement. Our results on two benchmark domains show that the joint architecture improves over the pipeline architecture as well as the previous state-of-the-art approach based on graphical modeling. Learned pruning gives twice the speedup over the joint architecture with negligible loss in accuracy. The cyclic architecture is faster than the joint approach and performs competitively with it.

#7.5 Robust Plackett-Luce Model for k-ary Crowdsourced Preferences

Bo Han, University of Technology Sydney
Yuangang Pan, University of Technology Sydney
Ivor W. Tsang, University of Technology Sydney

The aggregation of k-ary preferences is an emerging ranking problem, which plays an important role in several aspects of our daily life, such as ordinal peer grading and online product recommendation. At the same time, crowdsourcing has become a trendy way to provide a plethora of k-ary preferences for this ranking problem, due to convenient platforms and low costs. However,

k-ary preferences from crowdsourced workers are often noisy, which inevitably degenerates the performance of traditional aggregation models. To address this challenge, in this paper, we present a RObust PlAckett-Luce (ROPAL) model. Specifically, to ensure the robustness, ROPAL integrates the Plackett-Luce model with a denoising vector. Based on the Kendall-tau distance, this vector corrects k-ary crowdsourced preferences with a certain probability. In addition, we propose an online Bayesian inference to make ROPAL scalable to large-scale preferences. We conduct comprehensive experiments on simulated and real-world datasets. Empirical results on "massive synthetic" and "real-world" datasets show that ROPAL with online Bayesian inference achieves substantial improvements in robustness and noisy worker detection over current approaches.

#7.6 Distributed Multi-task Classification: A Decentralized Online Learning Approach

Chi Zhang, Nanyang Technological University
Peilin Zhao, Ant Financial
Shuji Hao, ASTAR
Yeng Chai Soh, Nanyang Technological University
Bu Sung Lee, Nanyang Technological University
Steven C.H. Hoi, Singapore Management University

Although dispersing one single task (STL) to distributed learning nodes has been intensively studied by the previous research, multi-task learning (MTL) on distributed networks is still an area that has not been fully exploited, especially under decentralized settings. The challenge lies in the fact that different tasks may have different optimal learning weights while communication through the distributed network forces all tasks to converge to an unique classifier. In this paper, we present a novel algorithm to overcome this challenge and enable learning multiple tasks simultaneously on a decentralized distributed network. Specifically, the learning framework can be separated into two phases: (i) multi-task information is shared within each node on the first phase(ii) communication between nodes then leads the whole network to converge to a common minimizer. Theoretical analysis indicates that our algorithm achieves a O(\sqrt(T)) regret bound when compared with the best classifier in hindsight, which is further validated by experiments on both synthetic and real-world datasets.

#7.7 Crowdsourcing with Unsure Option

Yao-Xiang Ding, Nanjing University
Zhi-Hua Zhou, Nanjing University

One of the fundamental problems in crowdsourcing is the trade-off between the number of the workers needed for high-accuracy aggregation and the budget to pay. For saving budget, it is important to ensure high quality of the crowd-sourced labels, hence the total cost on label collection will be reduced. Since the self-confidence of the workers often has a close relationship with their abilities, a possible way for quality control is to request the workers to return the labels only when they feel confident, by means of providing unsure option to them. On the other hand, allowing workers to choose unsure option also leads to the potential danger of budget waste. In this work, we propose the analysis towards understanding when providing the unsure option indeed leads to significant cost reduction, as well as how the confidence threshold is set. We also propose an online mechanism, which is alternative for threshold selection when the estimation of the crowd ability distribution is difficult.

#5.1: Learning Convolutional Neural Networks using Hybrid Orthogonal Projection and Estimation
#5.2: Limits of End-to-End Learning
#5.3: Locally Smoothed Neural Networks
#5.4: Deep Competitive Pathway Networks
#5.5: Scale-Invariant Recognition by Weight-Shared CNNs in Parallel
#5.6: Nested LSTMs
#5.7: Neural-Power: Predict and Deploy Energy-Efficient Convolutional Neural Networks

#6.1: Regret for Expected Improvement over the Best-Observed Value and Stopping Condition
#6.2: Distributionally Robust Groupwise Regularization Estimator
#6.3: Rate Optimal Estimation for High Dimensional Spatial Covariance Matrices
#6.4: A Quantum-Inspired Ensemble Method and Quantum-Inspired Forest Regressors
#6.5: On the Flatness of Loss Surface for Two-layered ReLU Networks
#6.6: A Covariance Matrix Adaptation Evolution Strategy for Direct Policy Search in Reproducing Kernel Hilbert Space

#7.1: Instance Specific Discriminative Modal Pursuit: A Serialized Approach
#7.2: Multi-view Clustering with Adaptively Learned Graph
#7.3: Learning Predictive Leading Indicators for Forecasting Time Series Systems with Unknown Clusters of Forecast Tasks
#7.4: Multi-Task Structured Prediction for Entity Analysis: Search-Based Learning Algorithms
#7.5: Robust Plackett-Luce Model for k-ary Crowdsourced Preferences
#7.6: Distributed Multi-task Classification: A Decentralized Online Learning Approach
#7.7: Crowdsourcing with Unsure Option

Welcome Reception

Wednesday, 15 November, 6:30 pm

Grand Ballroom B, Baekyang-Nuri (Building #130), Yonsei University

There will be a welcome reception for all participants at Grand Ballroom, Baekyang-Nuri inside Yonsei campus. Light meals and beverages will be served between 6:30 and 8 pm. Feel free to bring your friends along to enjoy the reception.

Conference Banquet

Thursday, 16 November, 7 pm

FRADIA Restaurant (Address: 121-9 Jamwon-dong, Seocho-gu, Seoul • Phone: 02-3477-0033)

Bus will leave at 6:20 pm.

The banquet is located at a restaurant that is floating on the water of the Han River. One of the longest rivers in Korea, the Han River flows across the center of Seoul, sometimes called the "Miracle on the Han River," referring to its economic growth. The Han River is most beautiful in the evenings, reflecting the lights from the surrounding skyscrapers, buildings, streets, and traffic. The sound of the river waves blend well with the city's nightscape. The banquet will feature entertainment from a Korean traditional performance. Buses will take you to the banquet from Yonsei University and return you to the hotel after the banquet.

Conference Lunch

Wednesday to Friday, 11:30 am to 1 pm

Cafeteria, Student Union (Building #207)

Conference attendees will have simple lunch meals with lunch coupon at the next building, Student Union.

Shilla Stay Mapo Hotel

Location: 3 km away from the venue

Address: 83 Mapo-daero, Mapo-gu, Seoul

Phone: 82 2 6979-9000

Website: http://www.shillastay.com/mapo

Nearby subway: Gongdeok Station (Line #5, 6, Airport Railroad), Gate #1

Airport Limousine Bus #: 6015, Bus stop: Gongdeok Station

Free shuttle bus schedule between Shilla Stay Mapo Hotel and Yonsei University

Day	Shilla Stay Mapo Hotel →Yonsei University	Yonsei University →Shilla Stay Mapo Hotel	Remarks
Wednesday, 15 November	7:30 am – 8 am	8 pm	After welcome reception
Thursday, 16 November	7:30 am – 8 am	9 pm	After conference banquet
Friday, 17 November	7:30 am – 8 am	7 pm	After closing ceremony

Amanti Hotel Seoul

Location: 2.15 km away from the venue

Address: 31, World Cup Buk-Ro Mapo-Gu, Seoul 04001, Korea

Phone: +82 2 334-3111

Website: https://www.hotelamanti.com

Nearby subway: Hongik University Station (Line #2, Airport Railroad Express: AREX), Gate #1

Airport Limousine Bus #: 6002, Bus stop: Hongik University Station

CasaVille Sinchon Residence

Location: 800 m away from the venue

Address: 15F, Seogang-ro, Mapo-gu, Seoul, Korea

Phone: +82 2 6220-4000

Website: http://www.casaville-shinchon.co.kr

Nearby subway: Sinchon Station (Line #2), Gate #7

Airport Limousine Bus #: 6002, Bus stop: Sinchon ohgeori or Hyundai Department Store

Ayan Acharya
Jamal Atif
Wei Bi
Jakramate Bootkrajang
Ruichu Cai
Mark Carman
Kian Ming Chai
Laiwan Chan
Kyriakos Chatzidimitriou
Chao Chen
Shang-Tse Chen
Wei Chen
Li Cheng
Sylvain Chevallier
Zhang Chi
Jaegul Choo
Wei Chu
Zhenwen Dai
Guillaume Dehaene
Hu Ding
Kien Do
Thanh Nghi Do
Vojtech Franc
Eibe Frank
Bin Fu
Wei Gao
Yuanjun Gao
Chen Gong
Vincent Guigue
Jiafeng Guo
Yandong Guo
Shaobo Han
Kohei Hatano
Di He
Jingrui He
Paul Honeine
Juhua Hu
Biwei Huang
Ding-Jiang Huang
Ruitong Huang
Viet Huynh
Tsuyoshi Ide
Cheongjae Jang
Wonseok Jeon
Wittawat Jitkrittum
Ioannis Katakis
Yiping Ke
Dongho Kim
Eun-Sol Kim
Geon-Hyeong Kim
Hyunwoo Kim
Jiseob Kim
Saehoon Kim
Minje Kim
Myunghwan Kim
Taehwan Kim
Arto Klami
Xiangnan Kong
Takuya Konishi
Tsung-Ting Kuo
Yanyan Lan
Hoai Minh Le
Tam Le
Trung Le
ByungJun Lee
Dongryeol Lee
Hung-yi Lee
Kanghoon Lee
Seunghak Lee
Juho Lee
Huitian Lei
Bin Li
Chun-Liang Li
Gang Li
Jiuyong Li
Ming Li
Limin Li
Shao-Yuan Li
Wen Li
Nan Li
Wenye Li
Xin Li
Lingqiao Liu
Liping Liu
Weiwei Liu
Wenyu Liu
Yashu Liu
Xinghua Lou
Chi-Jen Lu
Wei Lu
Ping Luo
Wei Luo
Yifei Ma
Qi Mao
Sanparith Marukatat
Sherin Mathews
Deyu Meng
Daichi Mochihashi
Jacob Montiel
Xin Mu
Atsuyoshi Nakamura
Tu Nguyen
Gang Niu
Jinoh Oh
Takuma Otsuka
Stefanos Ougiaroglou
Yingwei Pan
Sunho Park
Ioannis Partalas
Trang Pham
Duy Nhat Phan
Nico Piatkowski
Chao Qian
Sutharshan Rajasegarar
Maxime Sangnier
Bin Shen
Weiwei Shen
Kyung-Ah Sohn
Hang Su
Mahito Sugiyama
Ichiro Takeuchi
Koh Takeuchi
Ichigaku Takigawa
Ben Tan
Mingkui Tan
Cheng Tang
Khoat Than
Fei Tian
Ming-Feng Tsai
Shikui Tu
Xuan Thanh Vo
Chen Wang
Dayong Wang
Jialei Wang
Yang Wang
Li Wang
Wei Wang
Wei Wang
Yuyang Wang
Zheng Wang
Weiran Wang
Yu-Chiang Frank Wang
Yang Wang
Paul Weng
Qiang Wu
Feng Wu
Lei Wu
Shuo Xiang
Min Xiao
Chang Xu
Linli Xu
Chang Xu
Miao Xu
Zhixiang (Eddie) Xu
Zenglin Xu
Makoto Yamada
Shuang Yang
Can Yang
Haiqin Yang
Zhirong Yang
Hengshuai Yao
Yibo Yao
Nan Ye
Ian En-Hsu Yen
Sangwoong Yoon
Yuya Yoshikawa
Zhiwen Yu
De-Chuan Zhan
Kai Zhang
Chiyuan Zhang
Sen ZHANG
Yu Zhang
Lijun Zhang
Zongchang Zhang
Li Zhao
Shuigeng Zhou
Yue Zhu